Collaboratory for Multi-scale Chemical Science

http://cmcs.ca.sandia.gov/

Motivation & Vision

- Chemical Science is multi-scale and multi-disciplined and must be enabled by greater collaboration and information sharing.
- The Collaboratory for Multi-scale Chemical Science (CMCS) is an informatics infrastructure enabling collaborative synthesis of multi-scale information to create knowledge, piloted in the DOE chemical sciences community.

CMCS Team

Sandia National Laboratories

Larry Rahn	rahn@sandia.gov
Christine Yang	cyang@ca.sandia.gov
Carmen Pancerella	carmen@ca.sandia.gov
John Hewson	jchewso@ca.sandia.gov
Wendy Koegler	wkoegle@ca.sandia.gov
David Leahy	djleahy@sandia.gov
Michael Lee	mplee@ca.sandia.gov

Argonne National Laboratory

Al Wagner	wagner@tcg.anl.gov
Branko Ruscic	ruscic@anl.gov
Michael Minkoff	minkoff@mcs.anl.gov
Lee Liming	liming@mcs.anl.gov
Gregor von Laszewski	gregor@mcs.anl.gov
Sandra Bittner	bittner@mcs.anl.gov
Brian Moran	bmoran@anl.gov

Pacific Northwest National Laboratory

Jeff Nichols	jeff.nichols@pnl.gov
Brett Didier	brett.didier@pnl.gov
Theresa Windus	theresa.windus@pnl.gov
Jim Myers	jim.myers@pnl.gov
Karen Schuchardt	karen.schuchardt@pnl.gov
Eric Stephan	eric.stephan@pnl.gov

Lawrence Livermore National Laboratory

Bill Pitz	pitz1@llnl.gov

Los Alamos National Laboratory

David Montoya	dmont@lanl.gov

National Institute of Standards and Technology

Tom Allison	thomas.allison@nist.gov

Massachusetts Institute of Technology

Bill Green	whgreen@mit.edu

University of California, Berkeley

Michael Frenklach	myf@me.berkeley.edu

Goals

- **Develop an informatics infrastructure enabling dynamic addition of new workflows, schemas, scientific applications**
  - Portal enabling data-centric project- and community-level collaboration
  - Middleware and tools for security, notification, collaboration
  - Metadata-based data management technologies
  - Development environment for an evolving set of collaborative cross-scale science tools
- **Pilot informatics infrastructure within chemical science community, producing a new paradigm for distributed chemical science research**
  - Chemical science tools that generate, use and archive metadata
  - Develop collaborative data pedigree/annotation tools
  - Programmatic use of pedigree/annotation to provide cross-scale optimization and consistency checking

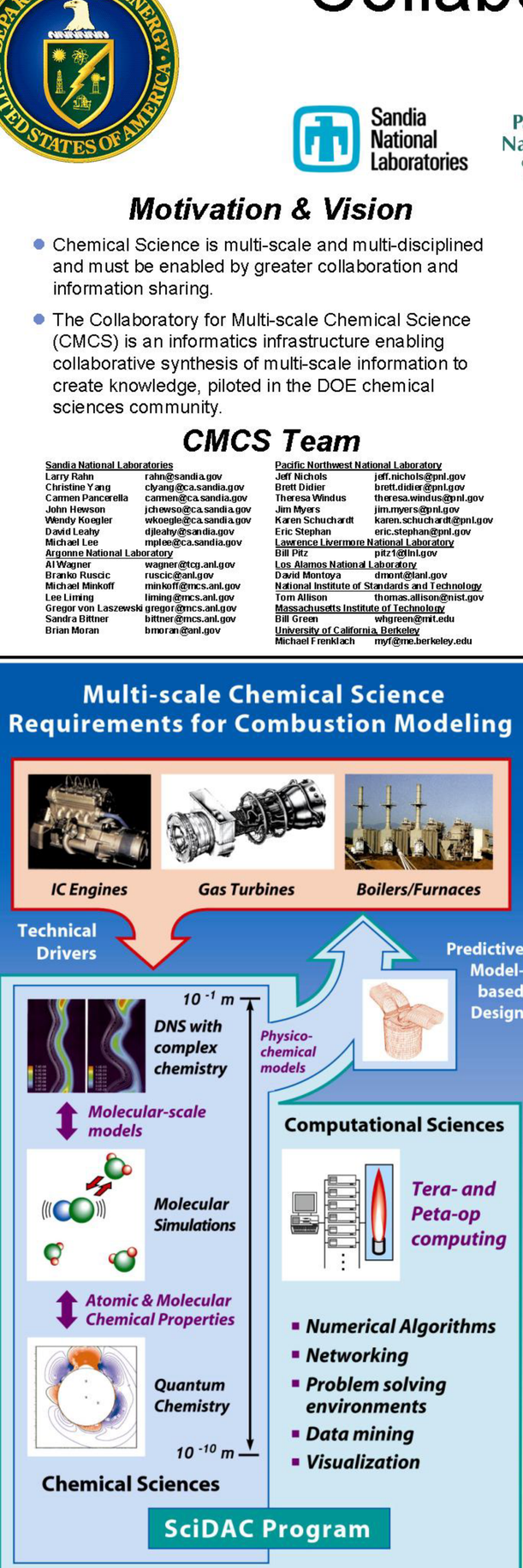
Challenges

- **Multi-scale chemical science information flow is complex.**
- **Facilitating collaboration and information transfer simultaneously across geography, scales, and disciplines (including computer and information sciences) is difficult.**
- **Making research meta-data and tools web-accessible invokes difficulties.**
  - Changes in behavior and assumptions of scientists.
  - Requires friendly, secure, easily deployed and maintained infrastructure.
  - May well invoke new problems in scientific community (pedigree/peer review/publication) and in industry/public.
- **Obtaining sufficient buy-in from users, institutions, and sponsors to maintain and evolve a production collaboratory infrastructure.**

Scientific Applications

- **Molecular Simulation**
  - Molecular Science Software Suite, MS3
  - NWChem/Ecce
- **Active Thermochemical Tables**
- **Databases and XML-based Data Standards at NIST**
- **GRI-Mech**
- **Feature Tracking & Detection**
- **Applications at the continuum scales**
  - Chemkin
  - Flamemaster
  - HCT
- **Reaction Mechanisms & Chemical Kinetics Models**
- **Reaction Rates**

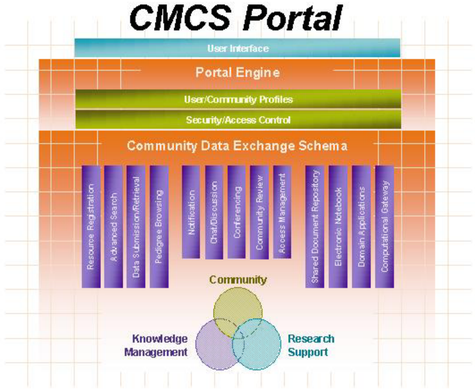
Multi-scale Chemical Science Requirements for Combustion Modeling



CMCS Architecture

Architecture diagram for CMCS showing portal integration of domain applications and data resources from across the multi-scale community.

CMCS Portal



Prototype design of portal for multi-scale chemical science.

Interactions With Other Projects

- **With CMCS Users/Customers**
  - Chemical science users and customers
    - Chemical scientists in DOE programs, laboratories
    - Research community, especially by enabling new projects
    - Chemical standards institutions and chemical applications industries
  - Computer, information & physical scientists
    - Piloting metadata/pedigree concepts of SAM project
    - Informatics infrastructure for other multi-scale science
  - DOE and other program managers
    - DOE/MICS NC program
    - DOE/BES chemical science program
    - Others enabled in the future; AFOSR, DOE/OIT, DOE/FE, …
- **With Other Projects**
  - Current & potential NC & SciDAC projects
    - Scientific Annotation Middleware (SAM)
    - DOE Science Grid
    - COG kits
    - Middleware technology to support science portals
    - Scientific data management ISIC
  - BES SciDAC computational projects
  - Other projects
    - OpenChem workbench – Tom McKinnon (Colorado School of Mines)
    - Bio/Spice – Adam Arkin (LBNL, UCB)
    - Science of Collaboratories – Gary Olson (University of Michigan)
    - Thermochemical models and databases - Mark Allendorf (SNL/CRF)

Infrastructure Technology

- **Data/metadata management**
  - Information sharing between people and applications across chemical scales. Incorporate third party applications and data stores. Dynamically evolve and extend schema.
- **Notification**
  - Notifivation of changes within CMCS without requiring user monitoring. Support both human and software centric types of 'workflow.'
- **Pedigree**
  - Adopt semantic standards. Consider pedigree to be a type of relationship.
- **Portal/personalization**
  - Mechanism to reduce information overload, provide filtering. Project context allows user to build/select from different sets of preferences.
- **Search**
  - Lower barriers to finding information that exists. Simplify adding new types of information.
- **Security**
  - Protect underlying resources, users' intellectual property. Assure accessibility, integrity, and persistence of data.